ParCoord Program User Guide

Thomas Boogaerts

13 June 2012

1 Quick Tutorial

When starting the program, a simple 5D dataset is loaded automatically. To get to know the program real quickly, follow these steps:

- Move the cursor over the red area. The lines over which the mouse hovers will be highlighted in black.
- To disable this highlighting, click on the black "HOVER button on the right of the screen once. Clicking a second time will turn the highlighting back on: the button is a toggle button (black means ON, grey means OFF).
- To make the red lines more transparent, leftclick on the textfield next to the red button on the right of the screen. The standard value for this textfield is 120. Remove the previous value by using the backspace button on the keyboard. Then type a desired opacity between 0 and 255, where 0 means completely transparent and 255 means completely opaque. Press enter.
- Now rightclick once on the red button. The red lines will become thicker. Right-clicking a second time will make the red lines disappear completely. Right-clicking a third time returns them to their initial state.
- Click on the green button on the left of the screen. This will cause the green group to become the "active" group. Now click on the "SL" toggle button (located between the "SE" and "SA" toggle buttons). Now click-hold-and-drag the mouse over some red area. A black line will appear. Release the mouse. This will cause all the red lines that intersect with the black line to turn green. These lines now belong to the green group.
- The "SA" button can be used to select lines according to the angle they make with a horizontal line. This works according to the same click-hold-and-drag principle: all the lines that have approximately the same slope as the black line will be selected. The range in which the angle can lie can be changed with the "SA Range" textfield.
- Load a new dataset by clicking on the "Load" button. Turn the toggle next to this button ON (black) when the labels are included in the dataset as first row. Turn the toggle next to this button OFF when this is not the case. Load labels for this dataset by clicking "Load labels". The file has to contain each label on a new line.

2 Full Function List with Annotations

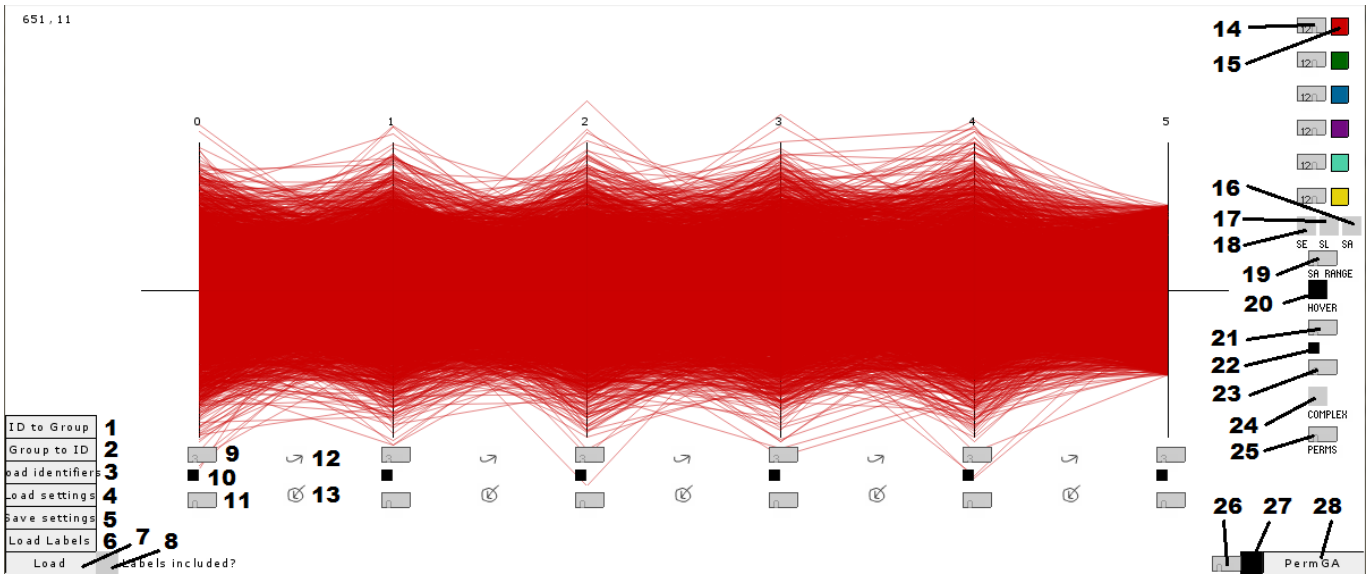


Figure 1: Main window of the *ParCoord* program (parallel coordinates plot)

The following list contains every function/button in the *ParCoord* program. The number corresponds to its position (see annotations in figure 1). Note that for toggle buttons, black means ON and grey means OFF. To change a value in a textfield, just click the textfield, delete the previous value using backspace, type the new value and press enter.

1. **ID to Group:** This button is used to load groups into the plot using identifiers. When each datapoint has an identifier associated with it (using the "Load Identifiers" button (nr. 3)), then a simple text file with on each line an identifier can be loaded, and all datapoints in the current plot that correspond to one of these identifiers in the text

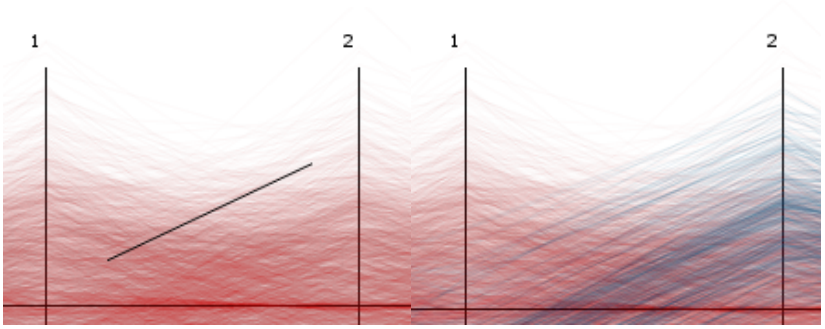
file will be associated with the currently active group. Note: the identifiers in the text file that have no corresponding datapoints in the plot will be ignored. It is important to first choose the active group in which you want the datapoints to be coloured (using the group buttons (nr. 15)). Also important: only one identifier per line in the text file.

2. **Group to ID:** This button does the opposite of button nr. 1. It simply saves a list of all the identifiers corresponding to the datapoints of the currently active group. Note that this button only works when each datapoints is associated with an identifier using the "Load Identifiers" button (nr. 3). Note that a non-empty active group has to be chosen (using buttons nr.15) in order to be able to save datapoints to a file, otherwise an empty file will be created.
3. **Load Identifiers:** This button simply associates an identifier with each datapoint in the dataset. Note that the list of identifiers needs to be a text file with an identifier on each line. The list has to have the same size as the number of datapoints in the dataset. The association is made in the original order of the dataset loaded, so for example the identifier on the first line of the text file is associated with the first entry in the dataset which was loaded into the *ParCoord* program. Note that loading a new set of identifiers overwrites previously defined identifiers.
4. **Load Settings:** All settings can be saved so that you can start working on a dataset and continue working at a later time. The settings are exported to a datafile and can be loaded back into the program by using this button. Note that you first have to load the database you were working on (by using the "Load" button (nr. 7)) before using the "Load Settings" button. Note that loaded settings will overwrite all current settings.
5. **Save Settings:** All settings can be saved so that you can start working on a dataset and continue working at a later time. The settings are exported to a datafile (you can choose the destination). The settings saved are the separation of the datapoints into groups, the opacities of the groups, the line thickness of the groups, the number of standard deviations displayed and the number of standard deviations the mean is shifted (in *SDs* mode) per axis separately, and if the axis is in *SDs* mode or in *MinMax* mode (see button nr. 9, 10 and 11).
6. **Load Labels:** This button can be used to load a label for each dimension in the dataset. The labels have to be in a text file, each line containing a separate label. Note that the number of labels has to correspond with the number of dimensions in the dataset. These labels will overwrite previously defined labels.
7. **Load:** This button is used to load a dataset into the program. Many formats of datasets can be loaded: most Delimiter Separated Values (DSV) files work, such as Comma Separated Values (CSV), Tab Separated Values (TSV), but also colon and semicolon separated values work. Infinity values can be denoted by "INFINITY", "Inf", "infinity", "INF" or "Infinity" and are plotted at 110% of the maximum value of the dimension they're in (for each dimension separately). Null values can be denoted by "NULL", "null", "Null", "none", "None", "" or "NaN". Null values are mostly ignored. Note that the toggle "Labels Included?" (button nr. 8) influences the load function. Discrete or class dimensions can be included in the dataset - meaning dimensions with only a few possible (non-numeric) values. These values will be plotted equidistantly.
8. **Labels Included?:** This toggle button causes the Load button (nr. 7) to use the first entry of the dataset as labels when it is switched ON and to just use the whole dataset as datapoints and use a numbering as labels when it is switched OFF. Note that labels can be loaded separately with the Load Labels button (nr. 6). When labels are included in the dataset but this button is switched OFF, the labels are treated as regular datapoints, causing all dimensions to become "discrete", meaning the values are treated as separate classes insteaf of a continuous spectrum of values. This considerably slows down the program and renders certain features useless. It is not recommended.
9. **Number of Standard Deviations to display in *SDs* mode:** This can be changed for each axis separately. In *SDs* mode (can be changed using button nr. 10), the data is displayed by using a set number of standard deviations to be displayed at each side of the mean. This number is three for each axis when first loading a dataset, which means that each black line representing a dimension has a length of six standard deviations (the length of a standard deviation is different for each dimension). Hovering over an axis in *SDs* mode shows the number of standard deviations and the original value of a virtual datapoint located at the position of the mouse cursor. Negative values just flip the data on the axis.
10. **Toggle between *SDs* mode and *MinMax* mode:** The mode can be changed for each axis separately. *SDs* mode displays a set number of standard deviations on each side of the mean (controlled by button nr. 9) using a certain number of standard deviations to shift the mean (controlled by button nr. 11). Note that, when changing button nr. 9, the virtual length of a standard deviation changes, which means the shift of the mean (nr. 11) is also influenced. Therefore it is most natural to first choose the number of standard deviations to display and then choose the mean offset (shifting of the mean). *MinMax* mode just plots the data so that the largest value is displayed at the top of the axis and the smallest value is displayed at the bottom of the axis. Note that Infinity values are displayed at the top as these are treated as 110% of the largest value. Buttons nr. 9 and 11 do not influence the axis when it is in *MinMax* mode.
11. **Number of Standard Deviations to shift the Position of the Mean in *SDs* mode:** When the number of standard deviations to display is chosen by button nr. 9 (which is essentially just choosing the virtual length of the standard deviation), the position of the mean value can be shifted a set distance. This distance is measured in numbers of standard deviations, *e.g.*, when the number of standard deviations to display for a specific axis is 3, then the length of one standard deviation in the plot is $\frac{1}{6}$ of the length of the axis. When the number of standard deviations to shift the position of the mean is changed to 1, then the data will be shifted to the top of the plot a distance of $\frac{1}{6}$ of the length of the axis.
12. **Switch Axes:** This button simply switches the dimensions (and accompanying buttons nr. 9, 10 and 11).

13. **Show Orthogonal Plot:** This button shows the orthogonal plot of the two dimensions it is located inbetween. This plot is displayed in the orthogonal plot window (see figure 6). The left axis is displayed as the x-axis in the orthogonal plot and the right axis is displayed as the y-axis.
14. **Change Opacity:** These buttons change the opacity for each group separately. Note that each group has an accompanying color. The standard opacity is 120, but can range from zero to 255, with zero being completely transparent and 255 being completely opaque. All the datapoints that are part of this group are displayed with the accompanying level of transparency/opacity, which means the number of lines crossing each other (the "density")and the groups to which these lines belong at each location of the plot can be estimated by looking at the darkness and color at this location.
15. **Select Active Group / Change Thickness:** Left clicking one of these group buttons will change the currently active group to the group you chose. This means the selection operators will insert all selected datapoints in this active group. The active group is also important for the ID to Group and Group to ID buttons (nr. 1 and 2). Right clicking one of these buttons once will change the line thickness for the corresponding datapoints from 1 pixel to 3 pixels. Right clicking a second time will make the lines of the datapoints belonging to this group disappear. This is handy using the selection operators to make more complex selections when the "Complex" toggle is switched ON (see button nr. 24). Right clicking a third time will set the line thickness back to 1 pixel.
16. ***SelectAngles* operator:** This toggle switches the *SelectAngles* mode ON or OFF. When this mode is switched ON, the *SelectLines* mode is switched OFF automatically. The *SelectAngles* mode enables the user to make a selection of datapoints by the angle the lines representing thsee datapoints between two dimensions make with a horizontal line, *e.g.*, when you want to select all the datapoints of which the lines representing them between dimensions X and Y make a 35° angle with a horizontal line. Note that a range of angles can be chosen using the "SA Range" textfield (nr. 19), *e.g.*, when the chosen angle is 35° and the "SA Range" textfield gives a 20° range, all datapoints with angles between 15° and 55° will be selected. The way the angle selection works is the following:

    (a) Select an active group.
    (b) Select a range of angles using the "SA Range" button (nr. 19).
    (c) To have a better overview of what's going on, turn off the "Hover" function (toggle button nr. 20).
    (d) Optionally change the group opacities using buttons nr. 14 to have a better overview.
    (e) Place the mouse cursor between the two dimensions in which the selection is to be made.
    (f) Leftclick-hold-and-drag the mouse, a black line will appear.
    (g) Move the mouse cursor (while still holding the left mouse button) until the black line has the same angle as the lines you want to select.
    (h) Release the left mouse button.

    In figure 2, an example of how to use the *SelectAngles* operator can be seen:



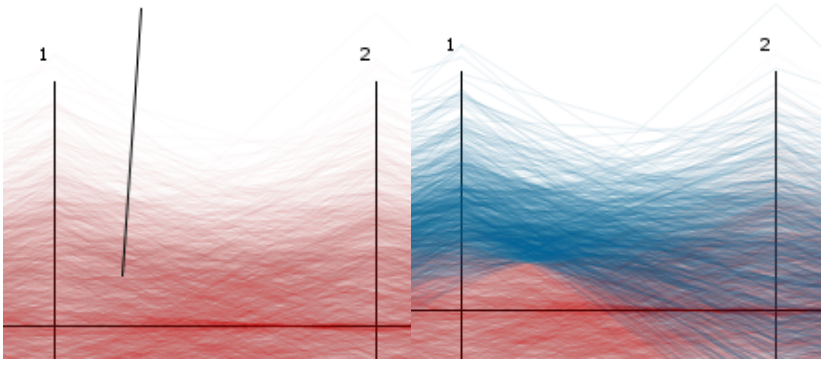
(a) Black Line created by leftclick-hold-and-drag of the mouse (b) Resulting selection in blue

Figure 2: Using the *SelectAngles* operator

17. ***SelectLines* operator:** This toggle switches the *SelectLines* mode ON or OFF. When this mode is switched ON, the *SelectAngles* mode is switched OFF automatically. This mode enables the user to make a selection of datapoints by the intersection of the polygonal lines representing these datapoints with a user-created line. This works as follows:

    (a) Select an active group.
    (b) To have a better overview of what's going on, turn off the "Hover" function (toggle button nr. 20).
    (c) Optionally change the group opacities using buttons nr. 14 to have a better overview.
    (d) Leftclick-hold-and-drag the mouse, a black line will appear.
    (e) All the polygonal lines (representing datapoints) that intersect with this black line will be selected and inserted in the active group.
    (f) Release the left mouse button.

In figure 3, an example of how to use the *SelectLines* operator can be seen:



(a) Black Line created by leftclick-hold-and-drag of the mouse

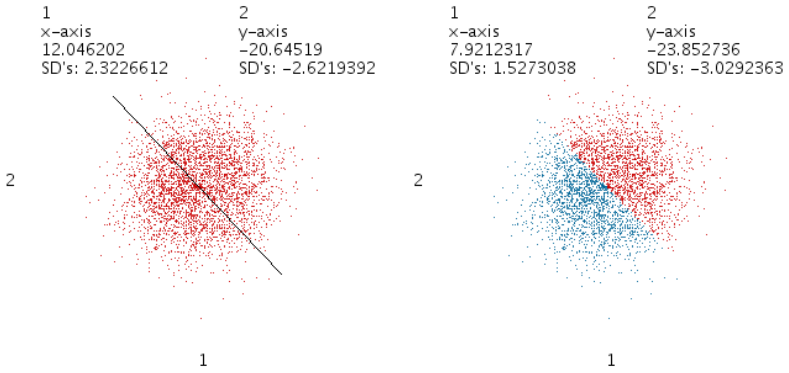
(b) Resulting selection in blue

Figure 3: Using the *SelectLines* operator

18. ***SelectOrthogonal* operator:** This toggle switches the *SelectOrthogonal* mode ON or OFF. When this mode is switched on, it is possible to select datapoints using the orthogonal plot (see figure 6). The selection is performed by the same click-and-drag principle as in the *SelectAngles* and *SelectLines* operator. This is how to use the operator:

    (a) Select an active group.
    (b) Optionally change the group opacities using buttons nr. 14 to have a better overview.
    (c) Choose a pair of dimensions in which the selection is to be made (possibly by manually rearranging the dimensions using button nr. 12), and display the orthogonal plot of this pair of dimensions by using button nr. 13.
    (d) Leftclick-hold-and-drag the mouse in the orthogonal plot, a black line will appear.
    (e) All the datapoints that are located to the right of the black line when facing the direction of the black line (from mouse-click to mouse-release) will be selected and inserted in the active group. Note that these are the datapoints located below the black line when the mouse-click point lies to the left of the mouse-release point.
    (f) Release the left mouse button.

    This can also be seen in figure 3:



(a) Black Line created by leftclick-hold-and-drag of the mouse

(b) Resulting selection in blue

Figure 4: Using the *SelectOrthogonal* operator

19. **SA Range:** This textfield is used to change the range of angles to select using the *SelectAngles* operator, as discussed in nr. 16.
20. **Hover:** This toggle button switches the highlighting ON or OFF when hovering over the polygonal lines. Note that datapoints that have null values are highlighted with endpoints (to the left and right of a null value) displayed as small black dots.
21. **Number of Standard Deviations to display in *SDs* mode for all axes:** This textfield simply performs the same function as nr. 9, but for all dimensions at once.
22. **Toggle between *SDs* mode and *MinMax* mode for all axes:** This textfield simply performs the same function as nr. 10, but for all dimensions at once.

23. **Number of Standard Deviations to shift the Position of the Mean in *SDs* mode for all axes:** This textfield simply performs the same function as nr. 11, but for all dimensions at once.
24. **Select Only out of Visible Lines ("Complex"):** When this toggle button is switched ON, the selection operators only operate on visible lines, which makes it possible to make more complex selection, *e.g.*, selecting a number of lines by using the *SelectLines* tool, and then out of the remaining lines selecting a number of lines that have a specified angle between dimensions X and Y.
25. **Calculate Permutations:** This textfield can be used to calculate a number of permutation orders of the dimensions so that the relations between all dimensions are visible. The number of permutations (and thus parallel coordinate plots) necessary to do this is $\lceil \frac{N}{2} \rceil$ for an N-dimensional dataset. Inserting values 1 to $\lceil \frac{N}{2} \rceil$ in the textfield results in the corresponding permutation orders to be displayed. Inserting value "0" in the textfield causes the original permutation order of the dataset to be displayed.
26. **Type of Genetic Algorithm:** There are four possible modes for the genetic algorithm, which can be run by inserting values 1-4 in this textfield and clicking the "Perm GA" button (nr. 28). The four modes (from 1 to 4) are:

    - **Mode 1**: The genetic algorithm finds a permutation order so that total positive correlation over all dimension pairs included in the permutation is maximized. The resulting permutation order strongly favours positive correlation, and a lot of highly positively correlated pairs are included in the plot. Note that no correlation is favoured over negative correlation in this case.
    - **Mode 2**: The genetic algorithm finds a permutation order so that total negative correlation over all dimension pairs included in the permutation is maximized. The resulting permutation order strongly favours negative correlation, and a lot of highly negatively correlated pairs are included in the plot. Note that no correlation is favoured over positive correlation in this case.
    - **Mode 3**: The genetic algorithm finds a permutation order so that total correlation (both negative and positive) over all dimension pairs included in the permutation is maximized. The resulting permutation order strongly favours correlation in general, and a lot of highly correlated pairs are included in the plot. Note that both positive and negative correlation are favoured over no correlation in this case.
    - **Mode 4**: The genetic algorithm finds a permutation order so that total correlation (both negative and positive) over all dimension pairs included in the permutation is minimized for the exact fitness function). Note that no correlation is favoured over both positive and negative correlation in this case.

27. **Pearson's or Spearman's:** The correlation measure used in the genetic algorithm is Pearson's when this button is toggled ON or Spearman's when this button is toggled OFF. Pearson's correlation favours linear correlation while Spearman's correlation is rank-based (which means it also takes nonlinear relations into account). Both correlation measures are rather similar.
28. **Run the Permutation order Genetic Algorithm:** This button makes the genetic algorithm run to find an optimal permutation order based on the settings of nr. 26 and 27.

Figure 5 contains an image of the density plot window. This window shows the histograms for each dimension and for each group separately, which is very useful for large datasets. Figure 6 shows an image of the orthogonal plot window. Figure 7 shows an image of both genetic algorithm windows. The first of these is the histograms window, which displays the number of individuals in the population with a certain fitness value for each fitness value. This is iterated for each generation, effectively creating a "moving histogram". The second window shows a plot of the mean fitness and best fitness of the population versus generation number.

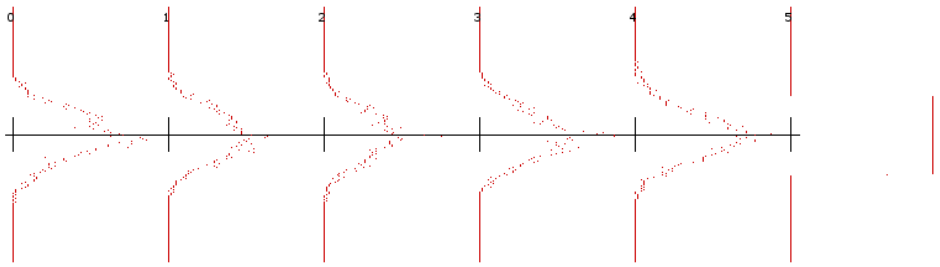


Figure 5: Density plots (histograms) window of the *ParCoord* program

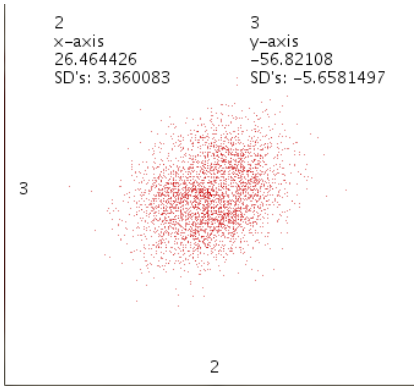
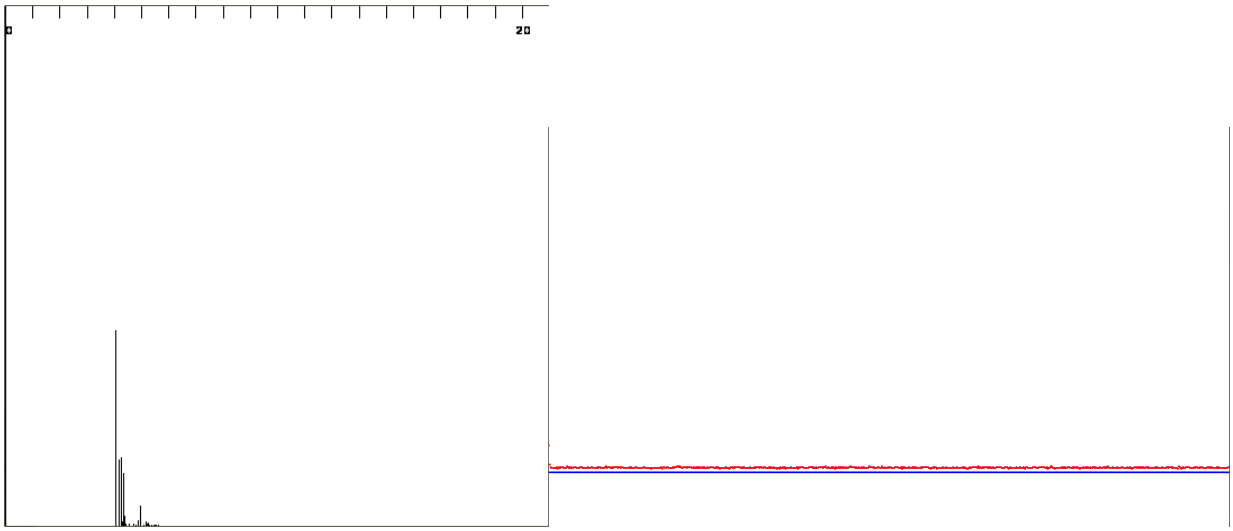


Figure 6: Orthogonal plot window of the *ParCoord* program



(a) Genetic algorithm Histogram (b) Mean and Best Fitness of Population versus Generation Number

Figure 7: Genetic algorithms windows of the *ParCoord* program